变分推断应用于教育测量模型

代立扬

华东师范大学调查与数据中心

摘要 变分推断常用于机器学习，是一种仅需编写少量代码便能实现较快计算速度的参数估计算法。研究展示了黑盒变分推断和均摊变分推断在教育测量模型上的应用，开发了可生成任意方差协方差矩阵的潜变量网络、可作为先验分布的神经相关矩阵，基于 sigmoid 或 softmax 的属性掌握模式网络等。实验结果显示变分推断在项目反应理论和部分 DINA 模型上的参数估计性能达到了顶尖水准。研究表明，变分推断能极大的帮助研究人员开发新的教育测量模型，也较适于普通用户在应用场景使用。

关键词 变分推断；项目反应理论；认知诊断模型；神经网络

1 引言

教育测量模型的研究与应用离不开参数估计算法的开发。项目反应理论（IRT）和认知诊断模型（CDM）是当今主流教育测量模型，常用的参数估计方法有 MCMC 算法、EM 算法等。

MCMC 是广泛应用于 IRT 和 CDM 的参数估计算法(Lu 等, 2018; Luo 等, 2017; Zhan 等, 2019)。MCMC 的优势是有 STAN(Carpenter 等, 2017)和 JAGS(Plummer, 2003)等软件供研究者使用，研究者只需编写少量代码便能应用 MCMC 进行参数估计，因此研究者常用 MCMC 开发参数较多、结构较复杂的模型，例如 HO-DINA 模型诞生之初便是应用 MCMC 进行参数估计(Torre 等, 2004)。MCMC 的缺点是计算耗时、内存消耗大，所以研究者开发新模型时需要拥有较大的耐心和较强的算力，而从用户的角度出发，MCMC 的这些缺点导致其难以应用于教育测量生产环境。

EM 算法是另一种常用的 IRT 和 CDM 参数估计算法(R. D. Bock 等, 1982; R. Darrell Bock 等, 1988; Torre, 2009, 2011)，并且在应用软件开发中取得了良好的效果(Chalmers, 2012; Chung 等, 2020; Ma 等, 2020)。比较有名气的 EM 算法是 R. D. Bock 等 (1982)开发的 BAEM，此外 EM 家族中还包括自适应积分算法(Schilling 等, 2005)，拉普拉斯近似算法(Huber 等, 2004)，

蒙特卡洛期望极大（MCEM）算法(Meng 等, 1996)等，这些算法的提出是为了处理高维潜变量 IRT 模型，但 Cai (2010)指出上述算法依然不适用于高维 IRT 模型，所以 Cai (2010)又开发了可以看作是随机 EM 的 MHRM。另外，虽然已有基于随机优化的 EM 算法应用于教育测量模型(S. L. Zhang 等, 2020)，但这些研究没有使用小批量优化（mini batch）和方差缩减技术，导致 EM 算法在大样本下的参数估计较为困难。最后， EM 算法对研究者的数学与代码功底有一定要求，所以研究者难以应用 EM 算法开发参数较多、结构较复杂的模型，典型案例是前面所提的 HO-DINA 模型，多年以后才有 EM 算法的参数估计实现形式(Ma 等, 2020)。

随着人工智能的发展，为解决大样本和多参数下贝叶斯模型的参数估计，计算机研究者发展了变分推断算法，并将该算法介绍给统计学研究者(Blei 等, 2017)，统计学研究者也开始使用变分推断(C. Zhang 等, 2019)。教育测量领域也有研究者建议使用变分推断进行参数估计(Linden, 2016)。随着概率编程软件的发展，变分推断的门槛不断降低，以至于可以像 MCMC 一样，研究者仅需编写少量代码，即可完成参数估计程序，易用性上超过 EM 算法，同时变分推断的运算时间少于 MCMC 算法(C. Zhang 等, 2019)，所以变分推断可能是非常适合教育测量模型的参数估计算法。

目前关于变分推断应用于教育测量模型的研究较少(Wu 等, 2020)，大部分涉及变分推断的教育测量模型研究主要集中于坐标上升变分推断(Cho 等, 2020; Hui 等, 2017; Imai 等, 2016; Yamaguchi, 2020; Yamaguchi 等, 2020)，基于机器学习思想发展而来且可应用概率编程软件的黑盒变分推断（black box variational inference）与均摊变分推断(amortized variational inference)的相关研究较少，仅有 IRT 参数估计的相关研究(Chen, 2017; Curi 等, 2019; Natesan 等, 2016; Wu 等, 2020)，CDM 方面暂时还是一片空白（Minka (2009)曾经以变分推断的名义展示 DINA 模型的参数估计，但实际使用的是期望传播）。坐标上升变分推断(Blei 等, 2017)需要研究者手动推导参数分布的期望解析式，有一定的技术门槛，且只适合可以求出参数分布期望解析式的模型，该算法通用性与易用性较差。黑盒变分推断与均摊变分推断可基于概率编程软件（pyro(Bingham 等, 2019)、edward(Tran 等, 2016)和 pymc3(Salvatier 等, 2016)等）编写代码，易用性较高。黑盒变分推断与均摊变分推断在教育测量研究中存在的问题主要有：一是损失函数梯度的选取，Wu 等 (2020)和 Curi 等 (2019)的梯度计算均是依据重参数算法(Kingma 等, 2014)，CDM 是否能应用重参数算法进行参数估计还亟待研究，此外，当前还缺乏应用 REINFORCE 梯度的研究，二是 Wu 等 (2020)和 Curi 等 (2019)的研究均是假设 IRT 潜变量的后验方差协方差矩阵是对角矩阵，而对角矩阵的信息量明显低于任意方差协方差矩

阵，三是 Wu 等 (2020)和 Curi 等 (2019)没有进行参数返真性实验，关于 IRT 参数的返真性问题还有待研究。

综上所述，变分推断是极具潜力的教育测量模型参数估计方法。本研究进一步展示黑盒变分推断和均摊变分推断在 IRT 的应用，并补上黑盒变分推断和均摊变分推断在 CDM 上的应用空白。

2 变分推断

2.1 损失函数

变分推断的原理是用简单的分布逼近复杂的真实分布。变分推断的损失函数称为 ELBO（Evidence Lower Bound），具体公式是

$$\mathrm{ELBO} \equiv \mathbb{E}_{q_\phi(\mathbf{z})}[\log p_\theta(\mathbf{x},\mathbf{z}) - \log q_\phi(\mathbf{z})]$$

这个公式由计算简单的分布与复杂的真实分布的 KL 距离而来，即

$$\mathrm{KL}(\mathrm{q}_\phi(\mathbf{z})\|\mathrm{p}_\theta(\mathbf{z}|\mathbf{x})) = \log p_\theta(\mathbf{x}) - \mathrm{ELBO}$$

上面的公式中，ϕ是变分参数，对于 IRT 来说，$\phi = \{\mu, \sigma\}$，μ是正态分布的 local 参数，σ是正态分布的 scale 参数，对于 CDM，$\phi = \{p\}$，p可以是伯努利分布的参数，也可以是分类分布(categorical distribution)的参数。

2.2 平均场

变分推断用简单的近似分布拟合真实分布，而简单的近似分布通常描述成平均场的形式，假设参数为ϕ_i，参数的分布为$q_i(\phi_i)$，则平均场为$q(\phi) = \prod_i q_i(\phi_i)$。平均场的优势：一是简化了近似分布的形式，二是可以利用动态图进行损失的反向传播(Wingate 等, 2013)。

2.3 坐标上升变分推断

Hui 等 (2017)和 Imai 等 (2016)应用坐标上升变分推断实现了 2 参数 probit 模型的参数估计，Yamaguchi (2020); Yamaguchi 等 (2020) 利用坐标上升变分推断实现对 DINA 模型和 MC-DINA 模型的参数估计。

坐标上升变分推断类似于 Gibbs 采样，即不断计算参数的期望，具体简略形式如下

$$\log(\phi_i(\theta_i)) \propto \mathrm{E}_{\phi_i \notin \phi}(\log(f(y|\phi)) + \log(p(\phi)))$$

坐标上升变分推断的缺点是难以推广到多参数 IRT，也难以推广到基于 logit 函数的 IRT（严格来说可以用近似的方式处理，但较为复杂(Ormerod 等, 2010)）。

2.4 黑盒变分推断

黑盒变分推断是 Ranganath 等 (2014)提出的算法。英语流利说(Chen, 2017)和 Natesan 等(2016)应用概率编程软件 edward 和 pyro 实现了一维 IRT 的黑盒变分推断参数估计。

黑盒变分推断的参数估计流程是，一是对变分参数ϕ求梯度得到

$$\nabla_\phi \mathrm{ELBO} = \mathbb{E}_{q_\phi(\mathbf{z})}\{[\nabla_\phi \log q_\phi(\mathbf{z})][\log \mathbf{p}(\mathbf{x},\mathbf{z}) - \log \mathbf{q}_\phi(\mathbf{z})]\}$$

从分布$q_\phi(z)$中采样z_s，$s = \{1\ 2\ 3\ \dots\ S\}$

于是得到变分参数的近似梯度

$$\nabla_\phi \mathrm{ELBO} = \frac{1}{S}\sum_{s=1}^{S}\{[\nabla_\phi \log q_\phi(z_s)][\log p(\mathbf{x},\mathbf{z_s}) - \log \mathbf{q}_\phi(\mathbf{z_s})]\}$$

然后更新变分参数

$$\phi = \phi + \rho\frac{1}{S}\sum_{s=1}^{S}\{[\nabla_\phi \log q_\phi(z_s)][\log p(\mathbf{x},\mathbf{z_s}) - \log \mathbf{q}_\phi(\mathbf{z_s})]\}$$

上式中ρ是 Robbins Monro 系数，主要应用于随机优化。

重复上述步骤直至变分参数ϕ收敛。

2.5 均摊变分推断

黑盒变分推断的劣势是大样本下潜变量参数量会爆炸，同时如果遇到新样本，黑盒变分推断需要重新学习，所以研究者提出了均摊变分推断(C. Zhang 等, 2019)。比较典型的均摊变分推断即变分自编码器(Kingma 等, 2014)，变分自编码器的思想是将神经网络作为分布参数的生成函数，这是利用了神经网络的万能逼近性质(Hornik 等, 1989)。应用均摊变分推断，近似分布$q_{\phi_i}(z_i)$改写为$q_{f(x_i)}(z_i)$。均摊变分推断常应用于图像生成等人工智能任务。

2.6 重参数化

教育测量模型即包含项目参数，又包含潜变量参数，所以计算梯度时会遇到下式

$$\nabla_{\theta,\phi} \mathrm{ELBO} = \nabla_{\theta,\phi}\mathbb{E}_{q_\phi(\mathbf{z})}[\log p_\theta(\mathbf{x},\mathbf{z}) - \log q_\phi(\mathbf{z})]$$

直接计算上式梯度较为困难，所以需要小技巧，例如重参数化。

正态分布的重参数方法，假设$\mathbf{z} \sim \mathcal{N}(\mu, \sigma)$，令$\epsilon \sim \mathcal{N}(0,1)$，则$\mathbf{z} = \epsilon * \sigma + \mu$，则ELOB的梯度为

$$\nabla_{\theta,\mu,\sigma}\text{ELBO} = \mathbb{E}_{q(\epsilon)}[\nabla_\theta \log p_\theta(\mathbf{x},\mathbf{z}) - \nabla_{\mu,\sigma}\log q_{\mu,\sigma}(\epsilon * \mu + \sigma)]$$

该重参数方法可应用于 IRT， VIBO 算法(Wu 等, 2020)便基于此重参数方法。

离散潜变量的重参数方法，存在一种叫 Gumbel Softmax 的算法(Jang 等, 2017)。Gumbel Softmax 的流程是假设二维向量v，从标准 Gumbel 分布中采样G_1,G_2，对应相加得到新的向量$v' = [v_1 + G_1, v_2 + G_2]$，通过 softmax 函数计算概率大小得到最终的类别，即$\sigma_\tau(v_i') = \frac{e^{v_i'/\tau}}{\sum_{j=1}^{n} e^{v_j'/\tau}}$。其中$\tau$是温度参数。CDM 理论上可以应用该方法的 HARD 模式（Straight Through Gumbel-Softmax）。

2.7 REINFORCE

REINFORCE 是另一种计算梯度的方法。REINFORCE 梯度的数学形式如下

$$\nabla_{\theta,\phi}\text{ELBO} = \mathbb{E}_{q_\phi(\mathbf{z})}\{\nabla_\phi \log q_\phi(\mathbf{z})[\log p_\theta(\mathbf{x},\mathbf{z}) - \log q_\phi(\mathbf{z})] + \nabla_\theta \log p_\theta(\mathbf{x},\mathbf{z}) - \nabla_\phi \log q_\phi(\mathbf{z})\}$$

该方法既可用于连续潜变量的梯度计算，也可用于离散潜变量的梯度计算。该方法在人工智能中常用于强化学习(Williams, 1992)，缺点是随机梯度的方差较大，但是可以应用方差缩减技术尽力克服该缺点。

2.8 随机优化与方差缩减

变分推断多应用于大规模数据集，所以研究者发展了随机优化和方差缩减的技术(Hoffman 等, 2013; Ranganath 等, 2014)，随机优化即每次从样本中采样小批量样本计算随机梯度。为了减少随机梯度的方差，研究者开发了包括 Rao-Blackwellization 和 Control Variates 等方差缩减方法(Ranganath 等, 2014)。

3 实验

3.1 模型

模型公式中参数的下标i代表样本序号，下标j代表项目（试题）序号，y_{ij}是输入数据（作答数据）。

3.1.1 项目反应理论

实验选取的模型是 2-4 参数 IRT 模型。

IRT 模型的数学形式如下。

$$P(y_{ij} \mid X_i) = c_j + \frac{d_j - c_j}{1 + exp(X_i \mathbf{a}_j + b_j)}, 0 < c_j < d_j < 1$$

上式中X_i是潜变量。

3.1.2 认知诊断模型

实验选取的模型是 DINA 模型和 HO-DINA 模型

DINA 模型的数学形式如下。

$$\mathrm{P}(\mathrm{y_{ij}}|\mathrm{\alpha_i}) = \mathrm{g_j}^{1-\eta_{ij}}(1-\mathrm{s_j})^{\eta_{ij}}, 0 < g_j < 1, 0 < s_j < 1, 0 < g_j + s_j < 1$$

或 Torre (2011)的重参数化版本

$$\mathrm{P}(y_{ij}|\alpha_i) = f_j + d_j\eta_{ij}, 0 < f_j < 1, 0 < f_j + d_j < 1$$

上式中$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$，$\alpha_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK}]$。

HO-DINA 模型的数学形式如下

$$\mathrm{P}(\alpha_{ik}|\theta_i) = \frac{\exp(\theta_i\lambda_{1k} + \lambda_{0k})}{1 + \exp(\theta_i\lambda_{1k} + \lambda_{0k})}, \alpha_{ik} \sim \mathcal{Bernoulli}(P(\alpha_{ik}|\theta_i))$$

上式中α_i是离散潜在变量（属性掌握模式），q_{jk}是Q矩阵的元素。

3.2 参数估计

实验代码基于 pyro 和 pytorch 编写，损失函数梯度基于 REINFORCE，仅在 CDM 中实验重参数梯度（基于 Gumbel Softmax 方法），方差缩减基于 Rao-Blackwellization 方法。实验

默认潜变量为随机参数，项目参数为确定参数，蒙特卡洛采样潜变量时只采样 1 个样本。

3.2.1 黑盒变分推断

一维 IRT 模型。潜变量的先验分布是$x_i \sim \mathcal{N}(0,1)$，潜变量后验分布是$x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$。该方法与 Wu 等 (2020)的方法一致。

多维 IRT 模型。潜变量的先验分布$X_i \sim \mathcal{N}(0, I)$。后验分布$X_i \sim \mathcal{N}(\mu_i, \Sigma_i)$，$\Sigma_i$是任意方差协方差矩阵。

DINA 模型。一是参考 Culpepper (2015)的 Gibbs 算法实现，$\boldsymbol{\alpha}_i = [\boldsymbol{\alpha}_{i0}, \boldsymbol{\alpha}_{i1}, \dots, \boldsymbol{\alpha}_{im}]$，$\boldsymbol{\alpha}_{i0} = [0,0,\dots,0]$，$\boldsymbol{\alpha}_{i0} = [1,0,\dots,0]$等，若属性掌握模式的维度为$K$，则$\boldsymbol{\alpha}_i$列表中包含$2^K$个元素，随机的从$\boldsymbol{\alpha}_i$中采样$\boldsymbol{\alpha}_{ih}$，$\boldsymbol{\alpha}_{ih} \sim \mathcal{categorical}([p_{i0}, p_{i1}, \dots, p_{im}])$，$\boldsymbol{\alpha}_{ih}$的先验分布是$\boldsymbol{\alpha}_{ih} \sim \mathcal{categorical}([1,1,\dots,1])$，该设定仅在低维属性掌握模式实验中使用；二是属性掌握模式的先验分布是$\alpha_{ik} \sim \mathcal{Bernoulli}(0.5)$，属性掌握模式的后验分布是$\alpha_{ik} \sim \mathcal{Bernoulli}(p_{ki})$，该设定仅在高维属性掌握模式实验中使用。

HO-DINA 模型。同样参考 Culpepper (2015)的 Gibbs 算法实现。高阶特质的先验分布是$\theta_i \sim \mathcal{N}(0,1)$，高阶特质后验分布是$\theta_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$。设$\boldsymbol{\alpha}_{ih} = [\alpha_{i1}, \alpha_{i2} \dots, \alpha_{ik}, \dots]$，则$p_{ih} = \prod \mathrm{P}(\alpha_{ik} | \theta_i)$，然后随机的从$\boldsymbol{\alpha}_i$的列表中采样属性掌握模式$\boldsymbol{\alpha}_{ih}$，$\boldsymbol{\alpha}_{ih} \sim \mathcal{categorical}([p_{i0}, p_{i1}, \dots, p_{im}])$。

3.2.2 均摊变分推断

一维 IRT 模型。与 Wu 等 (2020)方案一致，先验分布是标准正态分布，后验分布参数的生成器模型

$$h_i = g(y_i)$$

$$\mu_i = w_\mu h_i + b_\mu$$

$$\log \sigma_i{}^2 = W_\sigma h_i + b_\sigma$$

上式中g是激活函数。

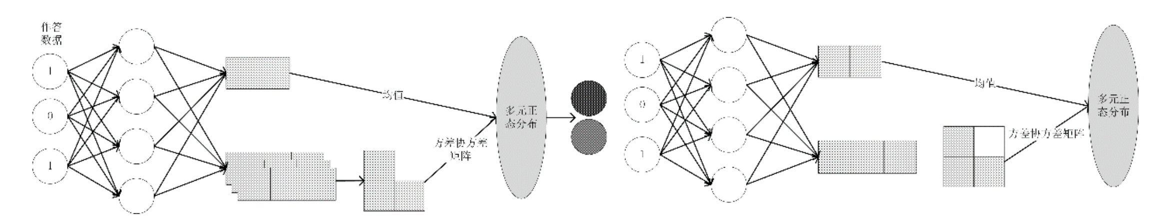


图 1 多维 IRT 潜变量生成器

多维 IRT 模型。后验分布 local 参数的生成器模型

$$h_i = g(y_i)$$

$$\mu_i = w_\mu h_i + b_\mu$$

后验分布方差协方差矩阵参数的生成器

$$\mathrm{tril}(L_i^*) = W_L h_i + b_L$$

$$L_i = \mathrm{tril_}(L_i^*) + \exp\left(\mathrm{diag}(L_i^*)\right)$$

$$\Sigma_i = L_i L_i^T$$

tril代表取矩阵的下三角元素，tril_代表取矩阵下三角元素且不包括对角线元素。

还有一种是潜变量共享方差协方差矩阵，生成器形式如下

$$\mathrm{tril}(L^*) = \frac{\sum \mathrm{tril}(L_i^*)}{N}$$

$$L = \mathrm{tril_}(L^*) + \exp(\mathrm{diag}(L^*))$$

$$\Sigma = LL^T$$

上式中N是 mini batch 的大小。

先验分布的设定，除了$X_i \sim \mathcal{N}(0, I)$，均摊变分推断还引入$X_i \sim \mathcal{N}(0, \Omega)$或$X_i \sim \mathcal{N}(0, \Omega_\mathrm{i})$，$\Omega$是相关矩阵，$\Omega$的计算与约束参照 stan 软件的方案(StanDevelopmentTeam, 2019)。使用神经网络计算的Ω，实验称为神经相关矩阵。

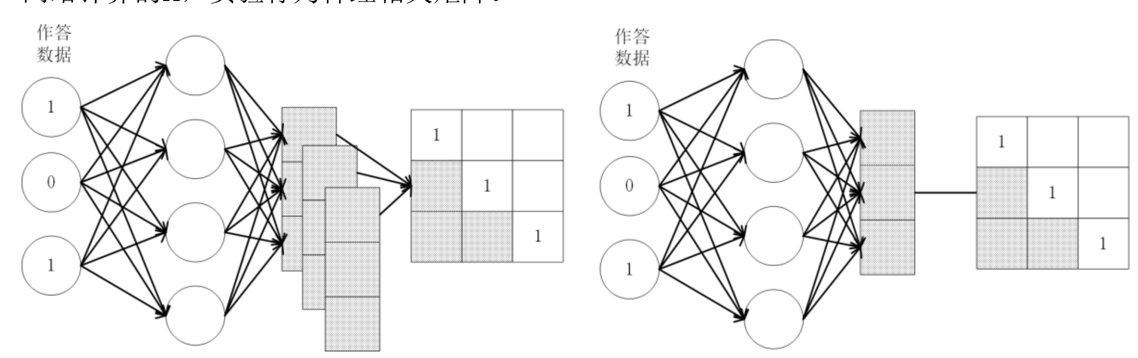


图 2 神经相关矩阵

该形式受离散变分自编码器的先验分布可以是受限玻尔兹曼机而启发。$X_i \sim \mathcal{N}(0, I)$等价于$L_2$惩罚，所以先验分布是对没有量纲的潜变量参数进行约束(Baker 等, 2004)，相关矩阵的对角线保证了潜变量参数的量纲约束，而相关系数的自由估计又保证了更多信息的纳入。

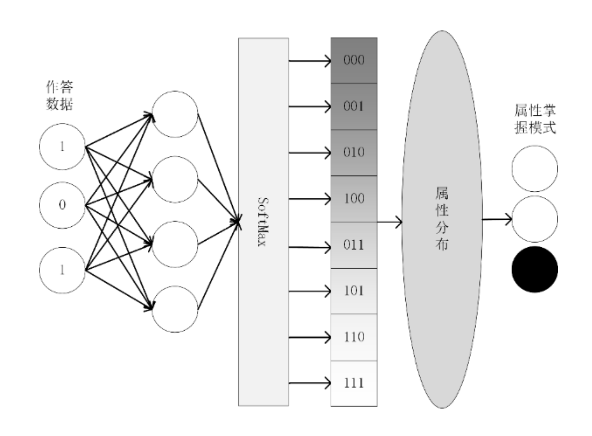


图 3 DINA 模型属性掌握模式生成器 1

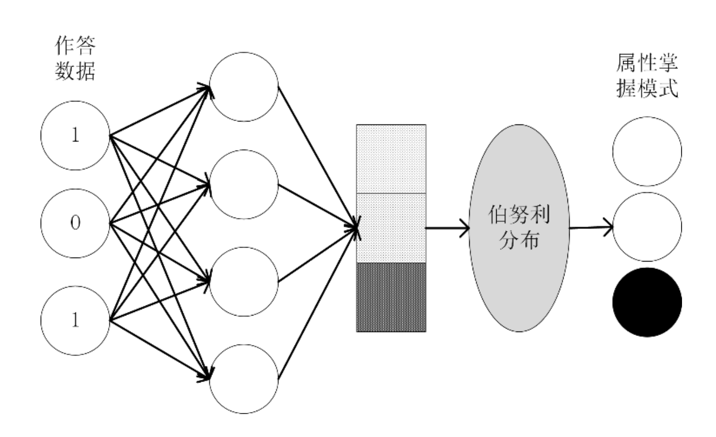


图 4 DINA 模型属性掌握模式生成器 2

DINA 模型。参考 Culpepper (2015)实现的后验分布生成器是

$$h_i = g(y_i)$$

$$p_{im} = \text{softmax}(h_i)$$

与黑盒变分推断第二种形式一致的后验分布生成器

$$h_i = g(y_i)$$

$$p_{ik} = \text{sigmoid}(h_i)$$

先验分布与黑盒变分推断一致。

HO-DINA 模型。后验分布生成器和先验分布与一维项目反应模型一致。

3.2.3 缺失数据

实验把所有缺失数据当作可以忽略的缺失数据。实验不删除存在缺失数据的样本，除非该样本所有数据为缺失数据。缺失数据带入神经网络计算时，将缺失数据赋值为-1。

3.2.4 模型识别

实验采用与 Fraser 等 (1988)一致的多维 IRT 模型识别方法，flexmirt 和 R 包 mirt 也采用此方法。

3.2.5 与 VIBO 的区别

VIBO(Wu 等, 2020)是另一种变分推断算法，现有的基于黑盒变分推断与均摊变分推断的教育测量模型参数估计算法均可以看作是 VIBO。

损失梯度。VIBO 基于重参数化，实验基于 REINFORCE（除 CDM 会应用 Gumbel Softmax 重参数化进行实验），VIBO 无法应用于 CDM，实验可以应用于 CDM。

多维 IRT 潜变量后验分布。VIBO 的方差协方差矩阵为对角矩阵，实验的方差协方差矩阵为任意矩阵。

多维 IRT 潜变量先验分布。VIBO 的方差协方差矩阵是单位矩阵，实验的方差协方差矩阵有单位矩阵和和神经相关矩阵。

缺失数据。VIBO 较为复杂，实验实现较为简单。

3.3 真实数据

3.3.1 LSAT

LSAT 是 Thissen (1982)发布的用于测试 IRT 模型的作答数据，来源于美国法学院入学委员会的法学院入学考试。LSAT 总共拥有 1000 样本，5 道试题。

3.3.2 PISA

PISA 是 OECD 进行的 15 岁学生阅读、数学、科学能力评价研究项目。实验选取的是 Wu 等 (2020)清洗后的 PISA 科学测试数据，二级计分，519334 个样本，183 题，其中有 73283 个全部数据为空的样本，去除空样本后，最终保留 446051 个样本，这些样本中包含 69014909 个缺失数据，总计缺失约 85%数据，仅有 15%的数据有效。

3.3.3 ECPE

ECPE 全称英语水平证书考试（examination for the certificate of proficiency in English），总共包含 2922 个样本，28 道试题，3 种属性，已在 Feng 等 (2014)等研究中使用过。

3.4 模拟数据

如不加特殊说明，IRT 模型的题量设定是 50，CDM 的题量设定是 100，每次实验模拟 10 份数据。

3.4.1 项目反应理论

$a_{kj} \sim unif(0.5,3)$，$b_j \sim \mathcal{N}(0,1)$，$c_j \sim unif(0.05,0.2)$，$d_j \sim unif(0.8,0.95)$，$a_{kj}$是第$k$维度上的斜率，$X_i \sim \mathcal{N}(0,I)$或$X_i \sim \mathcal{N}(0,\begin{bmatrix} 1 & corr \\ corr & 1 \end{bmatrix})$，$corr$取值 0.3,0.5 和 0.7。

3.4.2 认知诊断模型

第一种：$g_j \sim unif(0,0.3)$，$s_j \sim unif(0,0.3)$，$q_{kj} \sim bernoulli(0.5)$，$\alpha_{ik} \sim bernoulli(0.5)$，$\lambda_{i0} \sim \mathcal{N}(0,1)$，$\lambda_{i1} \sim unif(0.5,3)$，属性掌握模式的维度设定为 5。

第二种：参考 Li 等 (2020)的模拟方案，$g_j \sim unif(0,0.3)$, $s_j \sim unif(0,0.3)$, Q矩阵为$Q = \begin{bmatrix} I_K \\ Q_1 \\ Q_2 \end{bmatrix}$。$I_K$是$K$维单位矩阵。$Q_1 \in \{0,1\}_{K\times K}$，$Q_1$元素下标为$(i,i)$元素为 1，$i = 1,2,\ldots,K$，下标为$(i,i+1)$元素为 1，$i = 1,2,\ldots,K-1$，其余元素为 0。$Q_2 \in \{0,1\}_{K\times K}$，$Q_2$元素下标为$(i,i)$元素为 1，$i = 1,2,\ldots,K$，下标为$(i,i+1)$元素为 1，$i = 1,2,\ldots,K-1$，下标为$(i,i-1)$元素为 1，$i = 2,3,\ldots,K$，其余元素为 0。属性掌握模式的设定为$\theta_i = [\theta_{i1},\theta_{i2},\ldots\theta_{iK}]$，$\theta_i \sim \mathcal{N}(0,\Omega)$，$\Omega$是除了对角线元素为 1，其他元素为$\rho$的矩阵，实验设置$\rho = 0.3$。该设定仅用于实验维度较高的属性掌握模式。

3.4.3 缺失数据模拟

缺失数据模拟设定缺失 90%的数据。

3.5 对照软件及算法

IRT 实验的主要对照软件是 flexmirt 3.6.2(Chung 等, 2020)和 Wu 等 (2020)开发的 VIBO。实验对照算法是 BAEM、MHRM 和 VIBO。仅在四参数模型时用 R 包 MIRT 1.32.8 (Chalmers, 2012)代替 flexmirt。

CDM 实验的对照软件是 R 包 GDINA 2.8.0，对照算法是 EM。

3.6 硬件环境

实验环境的 CPU 是 Intel(R) Xeon(R) Gold 5217 3.00GHz，内存是 64G。

4 实验结果

4.1 真实数据

对于真实数据，仅使用均摊变分推断进行参数估计。评测指标使用 AUC (McClish, 1989)。所有的真实数据均拆分成测试集和验证集，数据拆分比例是 8:2。

4.1.1 项目反应理论

实验应用二参数 IRT 对数据进行分析。

表 1 LSAT 参数估计拟合统计量

先验分布	维度	测试集		验证集	
		AUC	边际似然	AUC	边际似然
单位矩阵	1 维	0.863	-2324	0.870	-586
	2 维	0.880	-2197	0.889	-555
	5 维	0.944	-2378	0.950	-597
神经相关矩阵	2 维	0.903	-2180	0.911	-549
	5 维	**1.0**	**-1292**	**1.0**	**-336**

表 2 PISA 参数估计拟合统计量

先验分布	维度	测试集		验证集	
		AUC	边际似然	AUC	边际似然
单位矩阵	1 维	0.832	-5607747	0.831	-1401557
	2 维	0.835	-5612647	0.835	-1402651
	5 维	0.835	-5625003	0.835	-1405852
神经相关矩阵	5 维	0.860	**-5343834**	0.860	**-1335769**
	10 维	0.894	**-4841089**	0.894	**-1211230**

表 1 和表 2 显示：（一）先验分布为神经相关矩阵比单位矩阵的拟合性能高；（二）高维模型比低维模型的拟合性能高。总体上，变分推断在 LSAT 和 PISA 上的拟合性能较为出色。

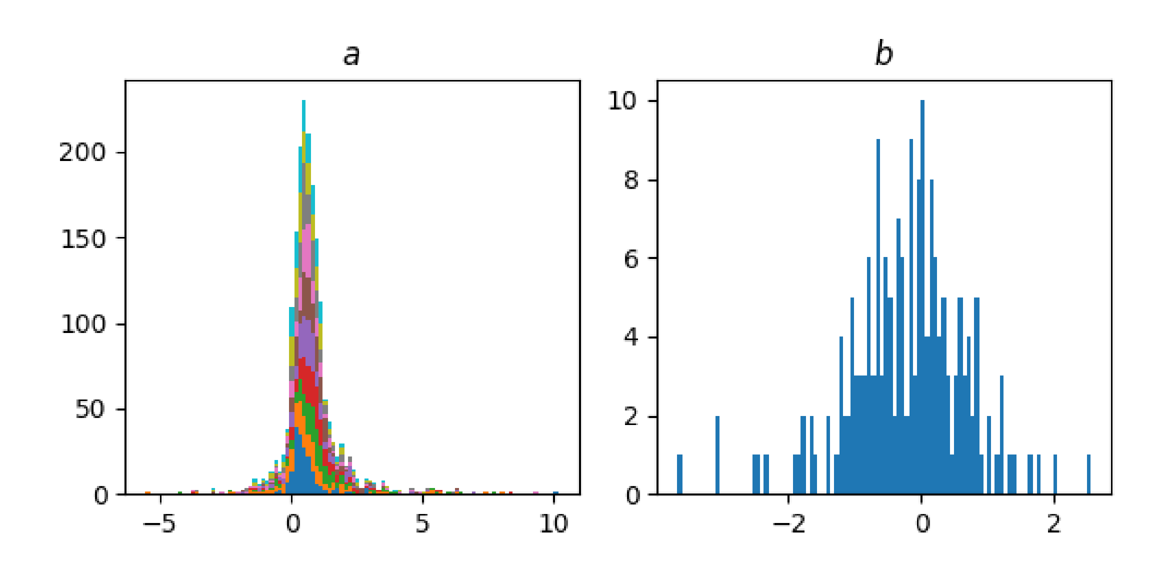


图 5 PISA 数据 10 维参数分布

4.1.2 认知诊断模型

表 3 ECPE 参数估计拟合统计量

模型	高阶特质维度	测试集		验证集	
		AUC	边际似然	AUC	边际似然
DINA	NA	**0.796**	**-31417**	**0.796**	**-7832**
HO-DINA	1 维	0.754	-40633	0.753	-10075

	2 维	0.746	-40752	0.745	-10128

表 3 显示：DINA 模型性能比 HO-DINA 较好。总体上，变分推断在 ECPE 上的拟合性能尚可。

4.2 模拟数据

实验结果中，BBVI 代表黑盒变分推断，AI 代表均摊变分推断，a、b、c、d、x代表 IRT 中相应的参数，g、s、λ_0、λ_1代表 CDM 中相应的参数。表格中的数字代表 RMSE 的平均数，括号内的数字代表 RMSE 的标准差。

4.2.1 项目反应理论

表 4 一维 IRT 参数估计返真 RMSE

样本	模型	算法	a	b	c	d	平均耗时（秒）
500	3 参数	BBVI	0.24(0.03)	0.24(0.03)	0.05(0.005)	NA	12
		AI	**0.24(0.02)**	**0.23(0.03)**	**0.05(0.004)**	NA	8
		BAEM	0.31(0.05)	0.33(0.04)	0.07(0.007)	NA	**3.5**
	4 参数	BBVI	0.38(0.03)	0.38(0.05)	**0.05(0.005)**	0.06(0.005)	16
		AI	**0.36(0.04)**	**0.35(0.04)**	0.05(0.006)	**0.04(0.005)**	12
		BAEM	0.43(0.06)	1.02(0.74)	0.4(0.05)	0.04(0.06)	**8**
1000	3 参数	BBVI	0.20(0.03)	0.19(0.03)	0.05(0.008)	NA	11
		AI	**0.19(0.02)**	**0.18(0.02)**	**0.05(0.007)**	NA	**8**
		BAEM	**0.19(0.02)**	0.25(0.04)	0.07(0.01)	NA	11
	4 参数	VI	0.37(0.04)	0.33(0.05)	0.05(0.005)	0.05(0.005)	24
		AI	**0.32(0.04)**	**0.28(0.03)**	0.05(0.005)	0.05(0.005)	17
		BAEM	0.35(0.03)	0.68(0.64)	**0.04(0.004)**	**0.04(0.003)**	**1**

表 5 多维 IRT 参数估计返真 RMSE

样本	模型	算法	a	b	平均耗时（秒）

1000	2 维 2 参数	BBVI	016(0.02)	**0.09(0.01)**	**24**
		AI	**0.14(0.02)**	**0.09(0.01)**	24
		VIBO	1.29(0.11)	0.11(0.03)	61
		MHRM	**0.13(0.01)**	0.13(0.01)	54
2000	3 维 2 参数	BBVI	0.17(0.04)	**0.06(0.01)**	95
		AI	**0.14(0.02)**	0.08(0.01)	**34**
		MHRM	0.13(0.01)	0.10(0.01)	126
3000	5 维 2 参数	BBVI	**0.25(0.03)**	**0.07(0.01)**	**115**
		AI	**0.28(0.01)**	**0.08(0.02)**	152
		MHRM	0.34(0.07)	0.12(0.07)	271

表 6 维度相关多维 IRT 参数估计返真 RMSE

模型	相关系数	算法	a	b	平均耗时（秒）
2 维 2 参数	0.3	AI	**0.09(0.03)**	**0.05(0.02)**	**15**
		MHRM	0.12(0.00)	0.08(0.00)	272
	0.5	AI	**0.16(0.02)**	**0.07(0.03)**	**16**
		MHRM	0.30(0.03)	0.08(0.03)	352
	0.7	AI	**0.22(0.05)**	**0.06(0.02)**	**99**
		MHRM	0.55(0.03)	0.11(0.03)	270
3 维 2 参数	0.3	AI	**0.13(0.04)**	**0.08(0.02)**	**99**
		MHRM	0..25(0.04)	0.08(0.05)	391
	0.5	AI	**0.19(0.03)**	**0.05(0.00)**	**99**
		MHRM	0.56(0.02)	0.06(0.02)	404
	0.7	AI	**0.36(0.01)**	**0.10(0.06)**	**145**
		MHRM	0.94(0.04)	0.16(0.04)	303

表 7 中等数据集 IRT 参数估计返真 RMSE

样本	题量	模型	算法	a	b	平均耗时（秒）
100k	1k	1 维 2 参数	AI	**0.09**	0.05	**1251**

	EM	0.41	**0.01**	1918
	MHRM	0.18	**0.01**	9878
3 维 2 参数	AI	**0.10**	0.07	**1450**
	MHRM	0.23	**0.05**	17429

表 4 未显示 MHRM，因为 MHRM 在一维多参数模型中表现不稳定（参数估计失败 2 次）；表 5 未显示 BAEM，因为 BAEM 在多维模型中过于耗时；表 6 中 flexmirt 软件设定潜变量协方差自由估计，变分推断使用神经相关矩阵作为先验分布。表 4、表 5、表 6 和表 7 显示：（1）维度相关模拟数据的参数返真，变分推断有较大优势且计算耗时低于 MHRM；（2）中等数据集的参数返真，变分推断有一定优势且计算耗时远低于 BAEM 和 MHRM；（3）一维多参数模型的参数返真，变分推断有一定优势，但计算耗时高于 BAEM；（4）多维模型的参数返真，变分推断与 MHRM 基本一致，计算耗时低于 MHRM；（5）VIBO 算法在二维模型中表现极其欠佳，证明了之前的猜测，VIBO 算法不适用多维 IRT 模型。

4.2.2 DINA 模型之重参数法 VS REINFORCE

表 8 重参数法 VS REINFORCE 梯度

样本量	梯度	g	s
1000	REINFORCE	**0.01**	**0.02**
	Gumbel Softmax	0.16	0.54

表 8 显示，Gumbel Softmax 重参数法难以应用于 DINA 模型，图 7 也验证了这个结果，图 6 的右图是属性掌握模式潜变量网络输出值与属性掌握模式真值的 AUC 值。

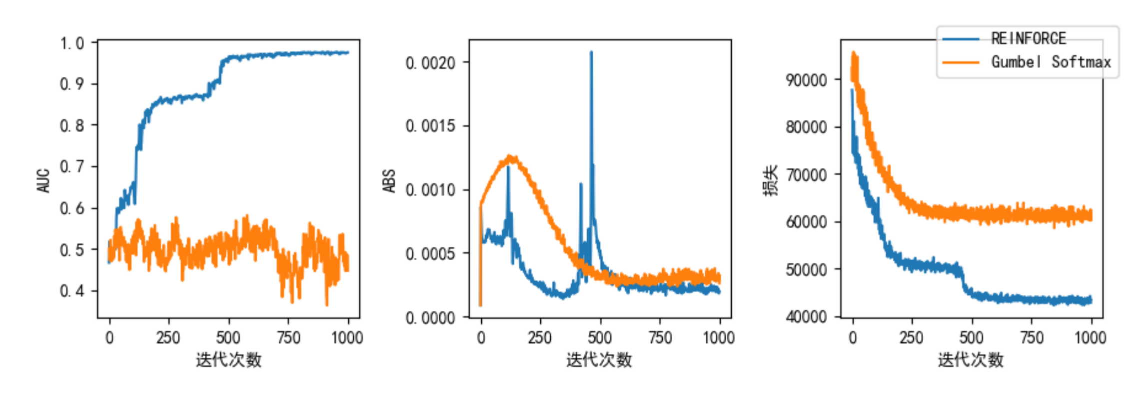


图 6 重参数法 VS REINFORCE 梯度

4.2.3 DINA 模型和 HO-DINA 模型

表 9 DINA 参数估计返真 RMSE

样本量	算法	g	s	平均耗时（秒）
500	BBVI	0.01(0.001)	0.03(0.005)	2
	AI	0.01(0.001)	0.03(0.005)	10
	EM	0.01(0.001)	**0.03(0.002)**	**0.2**
1000	BBVI	0.01(0.001)	0.02(0.005)	2
	AI	0.01(0.001)	0.02(0.005)	10
	EM	0.01(0.001)	**0.02(0.002)**	**0.4**
10000	AI	0.003(0.0003)	0.007(0.0009)	15
	EM	**0.003(0.0002)**	**0.006(0.0007)**	**3**

表 10 HO-DINA 参数估计返真 RMSE

模拟	算法	g	s	λ_0	λ_1	平均耗时（秒）
500	BBVI	0.01(0.001)	0.02(0.003)	0.14(0.02)	**0.26(0.11)**	8
	AI	0.01(0.001)	0.02(0.003)	**0.12(0.02)**	0.28(0.11)	11
	EM	0.01(0.001)	0.02(0.003)	0.15(0.02)	0.35(0.1)	**0.3**
1000	BBVI	0.01(0.000)	0.02(0.001)	**0.09(0.02)**	0.23(0.05)	36
	AI	0.01(0.000)	0.02(0.001)	0.09(0.03)	0.23(0.05)	31
	EM	0.01(0.000)	0.02(0.001)	0.09(0.03)	0.23(0.05)	0.5
10000	AI	0.004(0.000)	0.006(0.001)	0.05(0.02)	0.16(0.06)	15
	EM	**0.003(0.000)**	0.006(0.001)	**0.03(0.01)**	**0.06(0.03)**	4

表 9 和表 10 显示，变分推断在 CDM 模型上的参数估计返真精度与 EM 基本一致，但是计算耗时远高于 EM。

4.2.4 高维 DINA 模型

表 11 高维 DINA 模型参数估计返真 RMSE

样本量	属性掌握模式维度	g	s
1000	25	0.02	0.04
3000	50	0.02	0.06
10000	75	0.02	0.07

EM 在高维属性掌握模式 DINA 模型上的空间复杂度是$O(2^N)$，这导致 EM 难以应用于高维 DINA，所以实验另尝试了基于随机游走的 MCMC 算法，发现 MCMC 算法是变分推断的 3-10 倍（约耗时 3-10 小时不等），所以使用变分推断进行高维属性模式的参数估计是比较经济的选择。图 7 展示了高维属性掌握模式的返真程度。

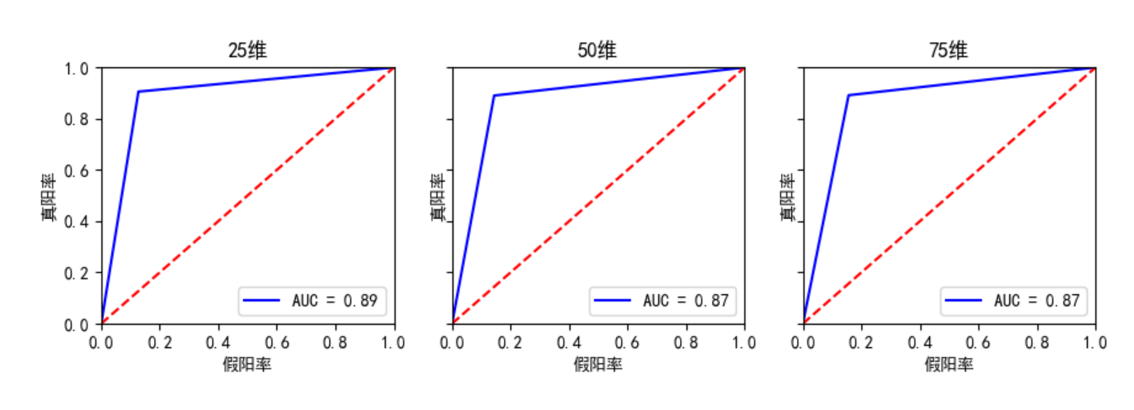


图 7 高维属性掌握模式 AUC

4.2.5 缺失数据

IRT 缺失数据实验题量设定为 500，模型为一维二参数，随机缺失 90%的数据。HO-DINA 缺失数据实验设定题量为 500，随机缺失 90%的数据。

表 12 IRT 缺失数据参数估计返真 RMSE

样本量	算法	a	b	平均耗时（秒）
1000	BBVI	0.20(0.007)	0.12(0.004)	120
	AI(500)	**0.17(0.005)**	0.12(0.004)	80
10000	AI(500)	0.09(0.009)	0.07(0.007)	100

表 13 HO-DINA 缺失数据参数估计返真 RMSE

样本量	算法	g	s	λ_0	λ_1	平均耗时（秒）
1000	BBVI	0.01(0.000)	0.03(0.003)	0.08(0.01)	0.22(0.03)	120
	AI	0.01(0.000)	0.03(0.003)	0.08(0.01)	0.22(0.03)	120
10000	AI	0.004(0.000)	0.006(0.001)	0.04(0.02)	0.20(0.07)	180

表 12 和表 13 显示，变分推断在处理缺失 90%的数据时依然能保持较好的参数返真性能。

5 结论与展望

无论是真实数据实验，还是模拟数据实验，均显示变分推断在教育测量模型上有较高的预测性能和参数返真性能。实验开发的潜变量网络模型以及神经相关矩阵，展示了变分推断的灵活性和扩展性，研究者可以利用神经网络的万能逼近性质或黑盒变分推断的简易性任意开发自己想要的模型。研究限于篇幅没有展示代码，但代码已开源，研究者通过开源代码可以发现编写变分推断的参数估计程序与编写 MCMC 的参数估计程序并无太多区别，变分推断非常有潜力应用于教育测量研究环境或实验环境，可以帮助研究者开发新的教育测量模型。

实验设计的算法在 IRT 领域达到了顶尖水准（the state of the art），在运行时间和参数返真上均领先或不弱于 flexmirt，所以极度推荐研究人员使用变分推断开发新的 IRT 模型，也推荐普通用户在应用场景中使用变分推断算法。实验设计的算法在 CDM 的表现差强人意，仅在高维属性掌握模式上超越 EM，尽管如此，还是推荐研究者使用变分推断开发新的 CDM 模型，毕竟变分推断能节省研究者的开发时间，至于普通用户，还是推荐使用 GDINA 实现的 EM 算法。

研究虽然显示了变分推断算法在教育测量领域拥有极大潜力，但还是存在一些问题，一是研究虽然给出了神经相关矩阵，并且通过模拟数据和真实数据实验获得良好效果，但是缺乏理论证明；二是实验设计的算法在 CDM 上的运行时间太长，这可能与实验采用的 REINFORCE 梯度有关系（REINFORCE 梯度的方差太大），而重参数方法在 CDM 上的使用不太理想，变分推断在 CDM 上的应用还需探索；三是标准化流（Normalizing Flow）的应用，潜变量的分布可能并不是简单的正态分布，所以可能需要标准化流来优化参数的分布，

实验其实测试了基于神经网络的自回归流，但是并未发现神经自回归流的优越性，所以并未在正文中阐述标准化流，这个有待后续研究；四是研究展示了变分推断的随机优化，其实 MCMC 和 EM 上也可以应用随机优化，即随机梯度 MCMC 以及随机 EM 算法，但是并未有相关研究在这两项参数估计技术应用于教育测量模型；最后，期望有更多的像变分推断这样的 AI 技术应用于教育领域。

参考文献

Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques*.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., . . . Goodman, N. D. (2019). Pyro: deep universal probabilistic programming. *Journal of Machine Learning Research, 20*(28), 973-978.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. *journal of the american statistical association, 112*(518), 859-877. doi:10.1080/01621459.2017.1285773

Bock, R. D., & Aitkin, M. (1982). Marginal maximum likelihood estimation of item parameters. *Psychometrika, 47*(3), 369-369. doi:10.1007/BF02294168

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-Information Item Factor Analysis. *applied psychological measurement, 12*(3), 261-280. doi:10.1177/014662168801200305

Cai, L. (2010). High-Dimensional Exploratory Item Factor Analysis by a Metropolis-Hastings Robbins-Monro Algorithm. *Psychometrika, 75*(1), 33-57. doi:DOI 10.1007/s11336-009-9136-x

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A Probabilistic Programming Language. *journal of statistical software, 76*(1), 1-32. doi:10.18637/JSS.V076.I01

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *journal of statistical software, 48*(1), 1-29. doi:10.18637/JSS.V048.I06

Chen, L. (2017). Fast Item Response Theory (IRT) Analysis by using GPUs. Retrieved from https://on-demand.gputechconf.com/gtcdc/2017/presentation/dc7176-lei-chen-fast-item-response-theory-irt-model-estimation-by-using-gpus.pdf

Cho, A. E., Wang, C., Zhang, X., & Xu, G. J. (2020). Gaussian variational estimation for multidimensional item response theory. *British Journal of Mathematical & Statistical Psychology*. Retrieved from <Go to ISI>://WOS:000577623600001

Chung, S., & Houts, C. (2020). flexMIRT: A Flexible Modeling Package for Multidimensional Item Response Models. *measurement interdisciplinary research & perspective, 18*(1), 40-54. doi:10.1080/15366367.2019.1693825

Culpepper, S. A. (2015). Bayesian Estimation of the DINA Model With Gibbs Sampling. *Journal of Educational and Behavioral Statistics, 40*(5), 454-476. Retrieved from <Go to ISI>://WOS:000363883900002

Curi, M., Converse, G. A., Hajewski, J., & Oliveira, S. (2019, 14-19 July 2019). *Interpretable Variational Autoencoders for Cognitive Models*. Paper presented at the 2019 International Joint Conference on Neural Networks (IJCNN).

Feng, Y., Habing, B. T., & Huebner, A. (2014). Parameter Estimation of the Reduced RUM Using the EM Algorithm. *applied psychological measurement, 38*(2), 137-150. doi:10.1177/0146621613502704

Fraser, C., & McDonald, R. P. (1988). NOHARM: Least Squares Item Factor Analysis. *multivariate behavioral research, 23*(2), 267-269. doi:10.1207/S15327906MBR2302_9

Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res., 14*(1), 1303–1347.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *neural networks, 2*(5), 359-366. doi:10.1016/0893-6080(89)90020-8

Huber, P., Ronchetti, E., & Victoria-Feser, M.-P. (2004). Estimation of Generalized Linear Latent Variable Models. *journal of the royal statistical society series b statistical methodology, 66*(4), 893-908. doi:10.1111/J.1467-9868.2004.05627.X

Hui, F. K. C., Warton, D. I., Ormerod, J. T., Haapaniemi, V., & Taskinen, S. (2017). Variational Approximations for Generalized Linear Latent Variable Models. *Journal of Computational and Graphical Statistics, 26*(1), 35-43. doi:10.1080/10618600.2016.1164708

Imai, K., Lo, J., & Olmsted, J. (2016). Fast Estimation of Ideal Points with Massive Data. *american political science review, 110*(4), 631-656. doi:10.1017/S000305541600037X

Jang, E., Gu, S., & Poole, B. (2017). *Categorical Reparameterization with Gumbel-Softmax*. Paper presented at the International Conference on Learning Representations.

Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes*. Paper presented at the International Conference on Learning Representations.

Li, C., Ma, C., & Xu, G. (2020). Learning Large Q-matrix by Restricted Boltzmann Machines. In.

Linden, W. J. v. d. (2016). Handbook of Item Response Theory, Volume Two: Statistical Tools. In.

Lu, J., Zhang, J., & Tao, J. (2018). Slice–Gibbs sampling algorithm for estimating the parameters of a multilevel item response model. *Journal of Mathematical Psychology, 82*, 12-25. doi:https://doi.org/10.1016/j.jmp.2017.10.005

Luo, Y., & Jiao, H. (2017). Using the Stan Program for Bayesian Item Response Theory. *Educational and Psychological Measurement, 78*(3), 384-408. doi:10.1177/0013164417693666

Ma, W., & Torre, J. d. l. (2020). GDINA: An R Package for Cognitive Diagnosis Modeling. *journal of statistical software, 93*(1), 1-26. doi:10.18637/JSS.V093.I14

McClish, D. K. (1989). Analyzing a portion of the ROC curve. *medical decision making, 9*(3), 190-195. doi:10.1177/0272989X8900900307

Meng, X.-L., & Schilling, S. (1996). Fitting Full-Information Item Factor Models and an Empirical Investigation of Bridge Sampling. *journal of the american statistical association, 91*(435), 1254-1267. doi:10.1080/01621459.1996.10476995

Minka, T. (2009). Automating Variational Inference for Statistics and Data Minin. *Invited talk at the 74th Annual Meeting of the Psychometric Society (IMPS 2009)*. Retrieved from https://www.microsoft.com/en-us/research/publication/automating-variational-inference-statistics-data-mining/

Natesan, P., Nandakumar, R., Minka, T., & Rubright, J. D. (2016). Bayesian Prior Choice in IRT Estimation Using MCMC and Variational Bayes. *Frontiers in Psychology, 7*. doi:ARTN 1422
10.3389/fpsyg.2016.01422

Ormerod, J. T., & Wand, M. P. (2010). Explaining Variational Approximations. *the american statistician, 64*(2), 140-153. doi:10.1198/TAST.2010.09058

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In.

Ranganath, R., Gerrish, S., & Blei, D. M. (2014). *Black Box Variational Inference*. Paper presented at the International Conference on Artificial Intelligence and Statistics.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic Programming in Python using PyMC3. *peerj, 2*. doi:10.7717/PEERJ-CS.55

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *psychometrika, 70*(3), 533-555. doi:10.1007/S11336-003-1141-X

StanDevelopmentTeam. (2019). Cholesky Factors of Correlation Matrices. In *Stan Reference Manual* (v2.18

ed.).
Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *psychometrika, 47*(2), 175-186. doi:10.1007/BF02296273
Torre, J. d. l. (2009). DINA Model and Parameter Estimation: A Didactic. *journal of educational and behavioral statistics, 34*(1), 115-130. doi:10.3102/1076998607309474
Torre, J. d. l. (2011). The Generalized DINA Model Framework. *psychometrika, 76*(2), 179-199. doi:10.1007/S11336-011-9207-7
Torre, J. D. L., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *psychometrika, 69*(3), 333-353. doi:10.1007/BF02295640
Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M. R., Liang, D., & Blei, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. In.
Williams, R. J. (1992). Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *machine learning, 8*(3), 229-256. doi:10.1007/BF00992696
Wingate, D., & Weber, T. (2013). Automated Variational Inference in Probabilistic Programming. In.
Wu, M., Davis, R. L., Domingue, B. W., Piech, C., & Goodman, N. D. (2020). Variational Item Response Theory: Fast, Accurate, and Expressive. In.
Yamaguchi, K. (2020). Variational Bayesian inference for the multiple-choice DINA model. *behaviormetrika, 47*(1), 159-187. doi:10.1007/S41237-020-00104-W
Yamaguchi, K., & Okada, K. (2020). Variational Bayes Inference for the DINA Model. *Journal of Educational and Behavioral Statistics, 45*(5), 569-597. doi:doi: 10.3102/1076998620911934
Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian Cognitive Diagnosis Modeling: A Tutorial. *Journal of Educational and Behavioral Statistics, Online First*. doi:10.3102/1076998619826040
Zhang, C., Bütepage, J., Kjellström, H., & Mandt, S. (2019). Advances in Variational Inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(8), 2008-2026. doi:10.1109/TPAMI.2018.2889774
Zhang, S. L., Chen, Y. X., & Liu, Y. (2020). An improved stochastic EM algorithm for large-scale full-information item factor analysis. *British Journal of Mathematical & Statistical Psychology, 73*(1), 44-71. Retrieved from <Go to ISI>://WOS:000509696500003